



Centraal Bureau voor de Statistiek

Divisie Methodologie en Kwaliteit
Sector Methodologie Heerlen

Steekproefontwerp NOPS en IS voor statistiekjaar 2009 plus bijlage met steekproeffracties

Thomas Kraan en Marc Smeets

In dit document worden het steekproefontwerp en de steekproeven van de Productiestatistiek in de nieuwe opzet en de Statistiek Investerings en Lease voor het statistiekjaar 2009 beschreven. Als bijlage is de tabel met de gespecificeerde steekproeffracties toegevoegd.

Trefwoorden: steekproefontwerp, Productiestatistiek, NOPS, Statistiek Investerings en Lease, enquêtevakantie, PS-plus.

1 Inleiding

In dit document worden het steekproefontwerp en de steekproeven beschreven voor de Nieuwe Opzet van de Productiestatistiek (NOPS) en de Statistiek Investerings en Lease (IS) voor het statistiekjaar 2009. Nieuwe steekproefontwerpen voor deze statistieken zijn gemaakt voor statistiekjaar 2007 (Smeets e.a., 2008) vanwege de invoering van de enquêtevakantie (in deze nota toegelicht) en voor statistiekjaar 2008 vanwege de overgang op de nieuwe bedrijvenindeling SBI2008 (Smeets e.a., 2009a). Er zijn drie redenen om voor statistiekjaar 2009 opnieuw naar de steekproef te kijken:

1. De belangrijkste aanleiding voor een nieuw steekproefontwerp is de nieuwe opzet van de Productiestatistiek (PS): NOPS (Aelen en Driesen, 2009). Het project NOPS is een onderdeel van het herontwerp keten van economische statistieken (HEcS), dat tevens o.m. de projecten Directe Ramingen Totalen (DRT) en Eenhedenbase (EHB) omvat. In deze nieuwe opzet worden de PS cijfers berekend uitgaande van DRT randen en worden responsdata gebruikt voor het schatten van de structuur (Aelen, 2009). Voorheen

Projectnummer:
BPA-nummer:
Datum:

DMH-207800
DMH-2010-04-16-TKRN
16-04-2010

werden de PS variabelen direct bepaald, uitgaande van responsdata en hulpvariabelen alleen.

2. Een andere belangrijke wijziging is de nieuwe definitie en indeling van de elementen van de populatie, vanaf nu bepaald door EHB. Dit betreft enerzijds de overgang op een andere statistische eenheid (het element van de steekproeftrekking), de zogenaamde ‘nieuwe bedrijfseenheid’ (aangeduid met BE_{nieuw}) en anderzijds het integraal waarnemen van de allergrootste ondernemingen (ConGO, Consistentie gegevens Grote Ondernemingen).
3. Voorts is er - voor een belangrijk deel gemotiveerd door de overgang op de BE_{nieuw} (zie 2.) - besloten door BES dat statistiekjaar 2009 een herijkingsjaar is voor de PS-plus. PS-plus strata zijn strata waar niet direct wordt waargenomen voor de PS maar waarvoor de stratumtotalen met hulpinformatie (BTW en vennootschapsbelasting) worden bepaald, onder gebruikmaking van structuren bepaald in eerdere jaren (zie Van de Burgt, 2006 en Stroeks, 2006). De herijking impliceert dat voor statistiekjaar 2009 er wel wordt waargenomen in de PS-plus strata, teneinde stratum- en kernceltotalen opnieuw direct te kunnen schatten en ook de structuren die gebruikt worden in toekomstige jaren opnieuw te bepalen.

Nadere informatie over deze wijzigingen en de gevolgen voor het ontwerp staat in het vervolg van deze nota.

Naast deze wijzigingen zijn veel randvoorwaarden voor het ontwerp voor statistiekjaar 2009 echter identiek aan die voor de ontwerpen voor de PS en de IS voor de statistiekjaren 2007 en 2008, i.h.b. de genoemde enquêtevakantie die de steekproefontwerpen voor de PS en de IS koppelt en (Smeets, e.a., 2008) en de stratificatie voor het steekproefontwerp naar SBI2008 (Smeets e.a., 2009a). Ook de kwaliteitscriteria zijn deels vergelijkbaar.

De invloed op het steekproefkader van de genoemde overgang op de nieuwe statistische eenheid, BE_{nieuw} , is beperkt (zie appendix B), met name voor de lagere grootteklassen, waarin het steekproefdeel van het steekproefontwerp zit. Met een beperkte populatiedynamiek is voor het steekproefdeel de populatie voor statistiekjaar 2009 qua eenheden en qua stratumaantallen daarom vergelijkbaar met die voor statistiekjaar 2008. Omdat een van de randvoorwaarden aan het steekproefontwerp voor statistiekjaar 2009 is dat deze geschikt is voor de oude ophoging (zoals in statistiekjaar 2008 en daarvoor) en dus voor statistiekjaar 2009 nog niet voor de nieuwe ophoging binnen NOPS ¹, kan, met de deels gelijke eisen als voorheen en vanwege de relatief kleine wijzigingen t.g.v. de overgang op BE_{nieuw} het steekproefontwerp goeddeels gelijk zijn aan dat voor statistiekjaar 2008. Daarom is het ontwerp voor statistiekjaar 2009

¹Voor het vervolg van NOPS wordt een steekproefontwerp voorzien dat wèl rekening houdt met de nieuwe ophoging binnen NOPS.

feitelijk een aanpassing van het ontwerp voor statistiekjaar 2008, gegeven de genoemde drie veranderingen. Deze nota licht de gevolgde werkwijze en de daarvoor noodzakelijke verificatie van de aannames toe.

Voor het overzicht van de lezer volgt een korte opsomming van de inhoud van deze nota. Een korte beschrijving van de PS en IS, hun populaties en de nieuwe kaderdefinitie, -indeling en -stratificatie staan in §2. De randvoorwaarden en uitgangspunten voor het steekproefontwerp en de globale werkwijze bij het bepalen van de steekproef staan in §3. Een uitleg van enquêtevakantie en de implementatie ervan voor statistiekjaar 2009 staat in §4. De kwaliteitscriteria waar het steekproefontwerp aan moet voldoen zijn te vinden in §5. Hoe deze zijn doorgerekend wordt uitgelegd in §6. De aanpassingen van het oude ontwerp leidend tot het ontwerp voor NOPS resp. IS voor statistiekjaar 2009 staan beschreven en gemotiveerd op basis van de kwaliteitsberekeningen in §7 resp. §8. Tot welke steekproeven dit heeft geleid staat in §9. De nota eindigt met enkele conclusies in §10. Appendix A bevat de tabel met de gespecificeerde steekproef fracties die uit het steekproefontwerp volgen. Appendix B is een inventarisatie van de relatie tussen het kader in termen van de nieuwe statistische eenheid en het kader in termen van de oude statistische eenheid.

2 De Productiestatistiek (NOPS) en de Statistiek Investerings en Lease (IS) en de veranderingen in kaderdefinitie en -indeling

In deze paragraaf wordt een korte beschrijving gegeven van de PS en de IS en de indeling van het kader.

2.1 Doel van de statistieken

De PS meet jaarlijks de productie van het Nederlandse bedrijfsleven en de IS meet jaarlijks de waarde van de duurzame productiemiddelen die de bedrijven in gebruik hebben genomen. Zie voor de PS Resing e.a. (2005), voor NOPS Lammertsma e.a. (2010) en voor de IS Brocke (2004).

2.2 De statistische eenheid

In eerdere jaren was de statistische eenheid voor de PS en de IS de zg. bedrijfseenheid (BE), die voortaan ‘oude bedrijfseenheid’ (BE_{oud}) genoemd zal worden. Met ingang van statistiekjaar 2009 is er een ‘nieuwe BE ’ (BE_{nieuw}) als gevolg van de overgang op de OG+. ² De OG+ verschilt van de oude OG op basis van welke zeggenschapsrelaties bepalend zijn voor de definitie

²OG staat voor ondernemingsgroep en een OG wordt bepaald op basis van zeggenschapsrelaties tussen ondernemingen. De plus (+) slaat op de veranderingen ontwikkeld binnen HEcS.

van een OG. Voor details wordt de lezer verwezen naar Beuken (2010). In de populatie wordt onderscheid gemaakt tussen de grootste (Top-X) en niet-Top-X bedrijven. Binnen de Top-X bedrijven valt de ConGO populatie, die bestaat uit de allergrootste ondernemingen. Voor de niet-Top-X ondernemingen geldt er een 1-op-1 relatie tussen OG+ en BE_{nieuw} . Voor de Top-X is er een relatie tussen OG+ en mogelijk meerdere BE_{nieuw} -en. De statistische eenheid, het element van de steekproeftrekking, is voor de hele populatie van NOPS en de IS de ‘nieuwe bedrijfseenheid’ (BE_{nieuw}).

De relatie tussen BE_{nieuw} en BE_{oud} is dat elke BE_{nieuw} uit één of meer BE_{oud} -en bestaat. In verreweg de meeste gevallen (96,4% voor een prototype van EHB) bestaat een BE_{nieuw} uit één BE_{oud} . Zie appendix B.

2.3 Indeling van het kader

De BE_{nieuw} -en zijn onderverdeeld in kerncellen en grootteklassen en kunnen wel of niet tot ConGO behoren. Een kerncel bestaat uit bedrijven die een SBI-code hebben die tot een bepaalde gedefinieerde deelverzameling van SBI-codes hoort. Elke kerncel correspondeert hierbij disjunct met een aantal SBI-codes. Voor statistiekjaar 2009 geldt de SBI- en kerncelindeling naar SBI2008.

De grootteklasse (gk) van een BE_{nieuw} is gebaseerd op het aantal werkzame personen en loopt van 0 t/m 9 (de grootteklasse op één digit is hier relevant). In het vervolg is alleen gk 1 t/m 9 van belang (op enkele uitzonderingen na, zie vervolg). Het kleinbedrijf, waarvoor enquêtevakantie wordt nagestreefd, omvat gk 1 t/m 3.

De driedelige indeling (kerncel, gk, ConGO/niet-ConGO) is tevens de stratificatie van het gestratificeerde steekproefontwerp. (De stratificatie naar ConGO/niet-ConGO is nieuw en hangt samen met de integrale waarneming binnen ConGO, zie §2.4). Een stratum wordt dus gekarakteriseerd met de labels kerncel, gk en het ConGO label. In een aantal gevallen zijn strata samengevoegd in gk 1 t/m 3 (als in eerdere jaren, zie Smeets e.a., 2008, en Smeets e.a., 2009a). De PS en de IS hebben dezelfde kerncellen, met dien verstande dat er voor sommige kerncellen niet voor beide statistieken PS en IS waarneming is.

2.4 Integrale waarneming binnen ConGO

Binnen de Top-X krijgen de allergrootste bedrijven, de ConGO ondernemingen, een speciale behandeling. De ConGO ondernemingen worden integraal, d.w.z. allemaal, waargenomen, zowel voor de PS als voor de IS. Om dit te bereiken worden deze apart behandeld van het steekproefdeel van de populatie door een extra stratificatie naar ConGO/niet-ConGO. Deze extra stratificatie geldt zowel voor de PS als voor de IS.

Zoals genoemd is er in alle ConGO strata integrale waarneming zowel voor de PS als voor de IS. Formeel is dit equivalent met een steekproef fractie identiek 1 in alle ConGO strata (zie boven). In de praktijk zijn de ConGO BE_{nieuw} -en echter handmatig toegevoegd aan de getrokken steekproef uit de niet-ConGO populatie. (Inclusief ConGO BE_{nieuw} -en uit gk 0 en mogelijk met SBI-codes die niet tot de PS of IS horen, om een compleet beeld voor ConGO te mogelijk te maken. Dit overeenkomstig gemaakte afspraken.)

2.5 Grootteklassen met integrale waarneming en grootteklassen met een steekproef

De PS en IS nemen waar in gk 1 t/m 9. Als in eerdere jaren neemt de PS bij afspraak integraal waar vanaf gk 6 en de IS vanaf gk 5. Dit bovenop de integrale waarneming voor ConGO voor alle gk-en. voor het niet-ConGO deel is er in de lagere grootteklassen, d.w.z. voor de PS gk 1 t/m 5 en voor de IS gk 1 t/m 4, een steekproef. Het kan (bijvoorbeeld in de industrie) voorkomen dat er ook in gk 5 voor de PS integrale waarneming is.

Een uitzondering op bovenstaande is dat in kerncel 32991 (sociale werkvoorziening) er wel in gk 0 wordt waargenomen (en wel integraal). Ook alle ConGO eenheden in gk 0 worden waargenomen.

3 Uitgangspunten voor het steekproefontwerp en aanpak ontwerp

3.1 Samenvatting uitgangspunten

De uitgangspunten en randvoorwaarden die gelden voor het steekproefontwerp luiden samengevat:

- Er worden steekproeven bepaald voor twee statistieken, de Productiestatistiek en de Statistiek Investerings en Lease, voor statistiekjaar 2009.
- De steekproeven voor beide statistieken worden getrokken volgens een gestratificeerd steekproefontwerp. De strata zijn hierbij gedefinieerd als kerncel maal grootteklasse maal ConGO/niet-ConGO. In de ConGO strata wordt integraal waargenomen. In de niet-ConGO strata wordt bij afspraak voor de PS in gk 6 t/m 9 integraal waargenomen en voor de IS in gk 5 t/m 9.
- De kerncelindeling waarvoor het steekproefontwerp gemaakt wordt is die volgens SBI2008 (conform Aelen en Driesen, 2009).
- Het steekproefontwerp wordt gemaakt aan de hand van afspraken en kwaliteitscriteria, geformuleerd door de projectleiding van NOPS en de IS en de senior users. Hierover meer in §5.

- De steekproefomvang van voor NOPS en de IS voor statistiekjaar 2009 zijn niet groter dan de steekproefomvang van voor statistiekjaar 2007 ³ voor de PS (80.549 enquêtes) en de IS (60.542 enquêtes). Voor de herijking van de PS-plus mogen 20.000 additionele enquêtes worden uitgezonden. Zie Aelen en Driesen (2009).
- Het steekproefontwerp voor NOPS hoeft niet aan te sluiten bij dat van eerdere jaren. Echter, later werd de voorkeur uitgesproken voor weinig verandering t.o.v. eerdere jaren. Er mag steekproefmassa verschoven worden tussen strata om beter aan de kwaliteitscriteria te kunnen voldoen. Met de projectleiding van de IS is afgesproken dat de steekproef zo veel mogelijk gelijk is aan die voor statistiekjaar 2008.
- Bij de bepaling van de steekproeven voor statistiekjaar 2009 wordt evenals voor statistiekjaar 2008 (Smeets e.a., 2009) rekening gehouden met de enquêtevakantie voor de PS en de IS. Zie §4.
- Bij het ontwerp wordt rekening gehouden met non-respons door bij het doorrekenen van de kwaliteitscriteria uit te gaan van de verwachte fractie bedrijven uit de populatie die in de respons zitten in plaats van de fractie bedrijven uit de populatie die in de steekproef zitten (de steekproeffractie). Door non-respons is eerstgenoemde fractie lager dan de tweede. De verwachte respons is gebaseerd op responspercentages uit eerdere jaren. Zie §6.2.
- Sinds 2004 worden de PS cijfers voor een deel van de strata niet via directe waarneming, maar met registerinformatie (BTW en vennootschapsbelasting) bepaald. Dit zijn de zogenaamde PS-plus strata (Van de Burgt, 2006, en Stroeks, 2006). In PS-plus strata was de steekproeffractie voor de PS in de afgelopen jaren derhalve nul. Er worden ook PS-plus strata binnen NOPS gedefinieerd. Er is besloten dat 2009 een herijkingsjaar is voor de PS-plus en dat er extra steekproefeenheden mogen worden aangewezen om de PS-plus te herijken. Daarom is voor statistiekjaar 2009 ook een allocatie voor de PS-plus strata gemaakt. De kwaliteitseisen voor PS-plus gerelateerde strata zijn gelijk aan die voor NOPS als geheel. In Stroeks, 2006, staat dat in een herijkingsjaar de steekproef zo ontworpen wordt dat een goede PS stand bepaald kan worden. Dit betekent dat voor de aan PS-plus strata gerelateerde kerncellen dezelfde kwaliteitseisen gelden als voor de overige kerncellen, zie §5. Het steekproefontwerp voor de herijking staat toegelicht in §7.3.

³Het statistiekjaar 2007 is als referentie genomen, omdat voor statistiekjaar 2008 de steekproef groter mocht zijn dan in 2007 omdat toen zowel voor de stratificatie naar SBI'93 als voor de stratificatie naar SBI2008 steekproeven moesten worden ontworpen. De steekproefomvang van bedroegen voor statistiekjaar 2008 85.737 eenheden voor de PS en 61.244 eenheden voor de IS.

- Er is een lijst van praktijkregels voor steekproeven voor de PS. Een van die regels vereist een minimale steekproefomvang van 7 bedrijven per stratum. Deze is echter ingegeven mede op basis van ophoging in IMPECT en de daarin bruikbare weegschema's en de daarbij horende nauwkeurigheid. Voor dit steekproefontwerp gelden echter de kwaliteitscriteria in termen van steekproefvarianties als leidraad, volgens de eisen zoals geformuleerd door de senior users. Daarom is deze eis dit jaar niet gehanteerd (zie ook §4). Voor die strata waar waarneming nodig is, is er wel een minimale steekproefomvang van 1.

3.2 Aanpak ontwerp

Vanwege de in §3.1 genoemde uitgangspunten kan de strategie voor het steekproefontwerp van eerdere jaren (Smeets e.a., 2008, en Smeets e.a., 2009), worden overgenomen. Er wordt uitgegaan van een uitgangsallocatie gebaseerd op de allocatie van het voorafgaande jaar, en vervolgens gekeken in hoeverre deze voldoet aan de geformuleerde kwaliteitseisen en andere wensen. Waar nodig worden dan fracties aangepast.

Zo is voor statistiekjaar 2009 als uitgangsallocatie de allocatie naar SBI2008 voor statistiekjaar 2008 in termen van fracties per stratum overgenomen en vervolgens geschikt gemaakt voor de nieuwe definitie van het kader, het integraal waarnemen van de ConGO strata, individuele aanpassingen op verzoek van de statistiekafdelingen voor bepaalde branches en de deels nieuwe kwaliteitseisen.

Deze werkwijze is mogelijk omdat de steekproef naar SBI2008 voor statistiekjaar 2008 al aan de eisen voor statistiekjaar 2008 voldeed (enquêtevakantie, nieuwe SBI2008 indeling en de kwaliteitseis per kerncel) en de wijzigingen voor het kader voor de lagere grootteklassen (waarin het steekproefdeel zich bevindt) beperkt zijn. Een groot deel van de inspanning was gericht op het verifiëren van de werkhypothese: het doorrekenen van de kwaliteitseisen van het minimaal aangepaste oude ontwerp in de nieuwe situatie. Vervolgens is aan de hand van de resultaten daarvan de definitieve allocatie bepaald door waar nodig steekproefmassa te verschuiven. Voor de PS-plus strata is wel een nieuwe allocatie bepaald, die past bij de algemene kwaliteitseisen voor statistiekjaar 2009. Dit conform gemaakte afspraken.

4 Enquêtevakantie

4.1 Definitie

Om tegemoet te komen aan de wens van verminderde administratieve lastendruk van bedrijven is in 2007 de enquêtevakantie ingevoerd voor bedrijven in het kleinbedrijf, gk 1 t/m 3 (Smeets e.a., 2008).

Enquêtevakantie wil zeggen dat een bedrijfseenheid in een jaar door slechts één van beide enquêtes PS en IS mag worden benaderd en de zekerheid krijgt ten minste één jaar niet geënuêteerd te worden als deze BE in het jaar ervoor in de steekproef van één van beide enquêtes zat.

Implementatie van enquêtevakantie impliceert een beperking op de fracties per stratum h , $f_{PS,h}$ en $f_{IS,h}$, voor PS resp. IS,

$$f_{PS,h} + f_{IS,h} \leq 0,45 \text{ voor de strata } h \text{ met gk 1, 2 of 3.} \quad (1)$$

Voor een motivering voor deze ongelijkheid zie §5.1 van Smeets e.a. (2009a), en §4.1 van Smeets e.a. (2009b).

4.2 Implementatie voor statistiekjaar 2009

4.2.1 Algemeen

Hoe de enquêtevakantie in eerdere jaren is geïmplementeerd is te lezen in Smeets e.a. (2008) en Smeets e.a. (2009a). Daar staan ook de uitzonderingen op enquêtevakantie vermeld. Relevant zijn de genoemde uitzonderingen voor de SBI2008, (Smeets e.a., 2009a, §5.3). Mede met het oog op de instabiliteit van het voorlopig kader (d.w.z. dat het sterk van het definitieve kader kon afwijken) op basis waarvan het ontwerp gemaakt moest worden, het grote politieke belang van de enquêtevakantie en de beperkte doorlooptijd van het project, zijn niet opnieuw uitzonderingen wat betreft enquêtevakantie vastgesteld: De toekenningen en uitzonderingen voor statistiekjaar 2008 zijn overgenomen. Er zijn echter wel extra uitzonderingen geïntroduceerd voor de herijking van de PS-plus (zie §7.3): de PS-plus strata zijn op eenmalige uitzondering gezet voor enquêtevakantie voor de PS, ook dit in overleg met projectleiding NOPS en de senior users. Dat betekent dat in de PS-plus strata voor statistiekjaar 2009 ongelijkheid (1) niet geldt, maar dat geldt $f_{IS,h} \leq 0,45$.

4.2.2 Behandeling historische enquêtevakantierechten bij de overgang van BE_{oud} op BE_{nieuw}

Voor de implementatie van de enquêtevakantie in statistiekjaar 2009 moeten enquêtevakanties toegekend worden aan BE_{nieuw} -en, terwijl ze in eerdere jaren golden voor BE_{oud} -en. Voor de vertaling van de enquêtevakantie is uitgegaan van de volgende regel, geldend voor een BE_{nieuw} :

als van een BE_{nieuw} de grootste samenstellende BE_{oud} recht heeft op enquêtevakantie, heeft die BE_{nieuw} enquêtevakantie, anders niet.

Hierdoor krijgt een BE_{oud} een ‘mogelijk (op zijn BE_{nieuw}) overdraagbaar recht op enquêtevakantie’. Een ‘mogelijk overdraagbaar recht op enquêtevakantie’ van

een BE_{oud} is afgelezen uit de steekproeven van de PS en IS van statistiekjaar 2008 en de toen geldende stratumindeling van de betreffende BE_{oud} : zat een BE_{oud} in een stratum waar enquêtevakantie gold en zat de BE_{oud} in de steekproef van vorig jaar, dan heeft de BE_{oud} een ‘mogelijk overdraagbaar recht op enquêtevakantie’ voor de BE_{nieuw} waar die bij hoort, anders niet. Het gaat om een ‘*mogelijk* overdraagbaar recht op enquêtevakantie’ van een BE_{oud} ; alleen als de BE_{oud} de grootste is van de BE_{nieuw} waar die bij hoort, resulteert het in een enquêtevakantie van die BE_{nieuw} .

Een BE_{oud} verliest dus geen ‘mogelijk overdraagbaar recht op enquêtevakantie’ doordat zijn stratumindeling verandert. Anders gezegd: een ‘mogelijk overdraagbaar recht op enquêtevakantie’ van een BE_{oud} moet opgebouwd worden, en oude rechten van BE_{oud} -en zijn behouden. (Dit laatste geldt niet bij populatiedynamiek naar PS-plus strata, waar dit jaar wel primaire waarneming is.)

Verreweg de meeste toekenningen van enquêtevakantie aan BE_{nieuw} -en zijn via een 1-op-1 relatie tussen die BE_{nieuw} en een BE_{oud} met enquêtevakantie tot stand gekomen en in slechts een kleine minderheid van gevallen speelde het boven geformuleerde algoritme een rol.

5 Kwaliteitscriteria

De criteria voor de PS steekproef zijn gedefinieerd in termen van de te verwachten kwaliteit, uitgedrukt in de variatiecoëfficiënt van de opgehoogde cijfers (totalen). De criteria zijn doorgerekend op verschillende aggregatieniveaus. Dit omdat de cijfers gebruikt worden zowel voor uitspraken over een branche (vanuit BES) als over de macro-economie (vanuit MNR). De door de senior users vastgestelde kwaliteitscriteria zijn voor de steekproef van NOPS voor statistiekjaar 2009:

- De variatiecoëfficiënt voor het totaal van de PS is kleiner dan 1%. Dit is een nieuwe eis voor NOPS.
- De variatiecoëfficiënt voor een REGKOL is kleiner dan 5%. REGKOL is een aggregatieniveau van de onderverdeling van de economische activiteiten die van belang is voor nationale rekeningen. Een REGKOL is in het algemeen wat groter dan een kerncel. Soms valt een REGKOL samen met een enkele kerncel, soms met een aantal kerncellen, soms ook met een beperkt aantal SBIs van een groepje kerncellen. Zie definitie REGKOL in literatuurlijst. De voorwaarde voor REGKOLs is een nieuwe eis voor NOPS.⁴

⁴De variatiecoëfficiënt op REGKOL niveau is in overleg alleen voor die REGKOLs doorgerekend die *volledig* samenvallen met één of meer kerncellen.

- De variatiecoëfficiënt voor een kerncel is kleiner dan 20%. De eis gold ook voor de statistiekjaren 2007 (Smeets e.a., 2008) en 2008 (Smeets e.a., 2009a).

De doelvariabelen waarnaar in dezen gekeken wordt zijn ‘omzet’ en ‘toegevoegde waarde’.

Voor de steekproef van de IS was de eis dat de steekproef zoveel mogelijk hetzelfde is als vorig jaar. In de praktijk zijn echter soms fracties aangepast wanneer de variatiecoëfficiënt voor de variabele ‘totaal van de investeringen’ voor een kerncel de 40% overschreed. Dit betekent een lagere kwaliteit dan voor de PS, maar dit is met de kleinere totale steekproefmassa en hogere (en lastiger te schatten) stratumvarianties voor de IS lastig te verhelpen. Hier komt bij dat voor de IS het hogere niveau van 3-digit SBI en niet ‘kerncel’ het publicatieniveau is.

Omdat onzeker is of de nieuwe ophoogmethode binnen NOPS dit jaar al bruikbaar zal zijn voor de bepaling van de PS cijfers, is op initiatief van de projectleiding NOPS bij de bepaling van het steekproefontwerp ervan uitgegaan dat de cijfers bepaald worden volgens de oude methode.

Dat betekent dat schattingen van de variatiecoëfficiënt net zo worden uitgevoerd als eerdere jaren (voor details Smeets e.a., 2008, Smeets e.a., 2009a). Ook voor de IS is de berekening identiek. Voor de PS zijn er wel aggregatieniveaus waarvoor de variatiecoëfficiënten in eerdere jaren niet werden uitgerekend (zie boven) en niet medebepalend waren.

In het vervolg worden de relevante aggregaten op de verschillende niveaus weergegeven met agg. Deze zijn gedefinieerd in termen van de samenstellende strata voor agg:

- alle strata in de hele PS voor de berekening van de variatiecoëfficiënt van het totaal van de PS.
- alle samenstellende strata die horen bij een REGKOL voor de berekening van de variatiecoëfficiënt per REGKOL.
- alle strata die horen bij een kerncel voor de berekening van een variatiecoëfficiënt per kerncel.

De variatiecoëfficiënten worden als volgt bepaald:

De totalen per aggregaat agg worden voor de PS uitgerekend als

$$\hat{Y}_{\text{agg}} = \sum_{h \in \text{agg}} N_h^{2009} \bar{y}_h \quad (2)$$

in termen van het geschatte stratumgemiddelde \bar{y}_h van de n_h responsdata $y_{h,k}$,

$$\bar{y}_h = \frac{1}{n_h} \sum_{k=1}^{n_h} y_{h,k} \quad (3)$$

en de stratum omvang N_h^{2009} . Voor de IS wordt een quotiëntschatter met als hulpvariabele aantal werkzame personen (WP) gebruikt, weergegeven met x ,

$$\hat{Y}_{\text{agg}} = \sum_{h \in \text{agg}} N_h^{2009} \bar{x}_h \frac{\bar{y}_h}{\bar{x}_h}. \quad (4)$$

Hier geldt dat \bar{x}_h het stratumgemiddelde van de hulpvariabele is.

De kwaliteit is gedefinieerd in termen van de variatiecoëfficiënt van de schatting van het totaal voor het aggregaat. Voor een variabele Y is deze variatiecoëfficiënt gedefinieerd als

$$C_V(\hat{Y}_{\text{agg}}) = \frac{\sqrt{V(\hat{Y}_{\text{agg}})}}{\hat{Y}_{\text{agg}}} \quad (5)$$

met daarin $V(\hat{Y}_{\text{agg}})$ de variantie van het totaal over het aggregaat. Deze is te schrijven als som over de som van de varianties binnen de samenstellende strata,

$$\hat{V}(\hat{Y}_{\text{agg}}) = \sum_{h \in \text{agg}} \frac{N_h^{2009} (N_h^{2009} - \hat{p}_h n_{PS,h}^{2009})}{\hat{p}_h n_{PS,h}^{2009}} \hat{S}_h^2. \quad (6)$$

agg is samengesteld uit de niet-ConGO strata in agg uit gk 1 t/m G_{maxstprf} ,⁵ met $G_{\text{maxsteekproef}}$ de grootste grootteklasse waarvoor er een steekproef is, gk 5 voor PS en gk 4 voor de IS. In de uitdrukkingen is \hat{p}_h een schatting van de respons in stratum h . Door deze mee te nemen bij het doorrekenen van de kwaliteitscriteria, wordt rekening gehouden met de non-respons.

De stratumvarianties voor de PS kunnen worden geschat als

$$\hat{S}_h^2 = \frac{1}{n_h - 1} \sum_{k \in h} (y_{h,k} - \bar{y}_h)^2 \quad (7)$$

en voor de IS als variantie van de quotiëntschatter,

$$\hat{S}_h^2 = \frac{1}{n_h - 1} \sum_{k \in h} \left(y_{h,k} - \frac{\bar{y}_h}{\bar{x}_h} x_{h,k} \right)^2, \quad (8)$$

waarin de aantallen werkzame personen van de respondenten (h, k) zijn weergegeven met $x_{h,k}$.

6 Doorrekenen kwaliteitscriteria steekproefontwerp NOPS en IS 2009

6.1 Gebruik van de responsdata

Om de kwaliteitscriteria door te rekenen is uitgegaan van responsdata van statistiekjaar 2007. Deze zijn beschikbaar per BE_{oud} . Voor het doorrekenen van de

⁵De strata waarvoor er integrale waarneming is, dus alle ConGO strata en van de niet-ConGO strata die voor de hogere grootteklassen, tellen dus niet mee bij de variantie. Voor het aggregaat-totaal tellen wèl alle strata van het aggregaat mee.

kwaliteitscriteria voor statistiekjaar 2009 is echter responsdata nodig in termen van de in statistiekjaar 2009 gebruikte statistische eenheid, BE_{nieuw} , nodig. Daarom moet responsdata voor statistiekjaar 2007 per BE_{nieuw} gesimuleerd worden, uitgaande van de respons in termen per BE_{oud} . Tennekes en Boonstra hebben hiervoor een algoritme opgesteld, waarvan de eigenschappen puntsgewijs luiden:

- Per BE_{nieuw} uit het kader van statistiekjaar 2009 is gekeken of één van de er aan gekoppelde BE_{oud} -en gerespondeerd heeft in 2007. Als dat het geval is, dan worden voor de er aan gekoppelde BE_{oud} -en die niet gerespondeerd hebben de doelvariabelen geïmputeerd op basis van een lineair model, met WP als hulpvariabele.
- Het lineaire model is gefit op de responderende BE_{oud} -en binnen de deel-populatie van BE_{oud} -en met een gelijke 3-digit SBI'93-code. In dit lineaire model wordt verondersteld dat de doelvariabele voor een BE_{oud} evenredig is met het aantal werkzame personen (WP). Per 3-digit SBI'93 geldt een aparte evenredigheidsfactor.
- Voor elke BE_{oud} is indien mogelijk de WP waarde voor statistiekjaar 2007 gebruikt, als de BE toen niet bestond de WP waarde voor 2008 en anders de WP waarde voor statistiekjaar 2009. (NB: de beschouwde BE_{oud} -en komen uit het (voorlopig) kader van 2009 en kunnen dus mogelijk in eerdere jaren nog niet bestaan.)
- Als alle WP waarden ongelijk 0 waren, is in het lineaire model een weegfactor $1/WP$ gebruikt, anders een weegfactor 1.
- Deze imputatie met een lineair model is alleen toegepast wanneer er 5 responderende BE_{oud} waren per 3 digit SBI'93. Zo ontstaat een set van BE_{nieuw} -en waarvoor alle er aan gekoppelde BE_{oud} -en een waarde van de doelvariabele is, deels op basis van empirie, deels op basis van imputatie met een lineair model met empirie als input. Door deze set per BE_{nieuw} te sommeren over de samenstellende BE_{oud} -en wordt een set gereconstrueerde responsdata op BE_{nieuw} niveau verkregen. Zo kan voor een groot deel van de BE_{nieuw} -en de deels waargenomen en deels geïmputeerde informatie op BE_{oud} niveau geaggregeerd worden tot een gereconstrueerde respons op BE_{nieuw} niveau.
- Bovenstaande imputatie lukt niet voor een BE_{nieuw} die deels is opgebouwd uit BE_{oud} 's die niet in het PS domein vallen (vb. een deel van een bedrijf dat onderwijs of cursussen verzorgt.) Dan is per BE_{nieuw} op BE_{oud} niveau een responswaarde gesimuleerd via een lineaire fit met WP aan de responswaarden voor die BE_{oud} -en waarvoor wel informatie beschikbaar is.

Uiteindelijk ontstaat een set met BE_{nieuw} -en met gereconstrueerde responsdata. Elke BE_{nieuw} staat in het door ons gebruikte kader met een SBI2008 code en gk. Met de lijst van kerncellen is daarmee aan elk element van de gereconstrueerde set responsdata op BE_{nieuw} niveau een stratum naar SBI2008 toe te kennen. Dan kan per SBI2008 stratum een stratumvariantie \hat{S}_h^2 geschat worden volgens de uitdrukkingen (7) en (8), waarmee de variatiecoëfficiënten op de verschillende aggregatieniveaus bepaald kunnen worden, via de uitdrukkingen (i.h.b. vgl. (5)) en definities in §5

6.2 Responsfracties

De gehanteerde responsfracties \hat{p}_h voor de BE_{nieuw} zijn de responsfracties in termen voor de BE_{oud} zoals die bij het steekproefontwerp voor de PS/IS2008 zijn gehanteerd. Dit in overleg met de projectleiding NOPS.

Deze keuze is te motiveren op basis van het grote percentage BE_{nieuw} -en die uit slechts één BE_{oud} bestaan, met name in het steekproefdeel van de populatie van de beide statistieken. Zie appendix B.

6.3 Variantieberekeningen

De stratumgemiddelden en -varianties zijn berekend als uitgelegd in §5 en §6.1. Echter, niet in alle strata was er voldoende gesimuleerde responsdata om een waarde te kunnen bepalen. Daarom zijn -in de gevallen waar dit nodig was - strata samengevoegd om varianties te berekenen, zoals dit ook bij steekproefontwerpen voor eerdere statistiekjaren gebeurd is.

- Voor zowel de PS als de IS is als in één van de grootteklassen 1, 2, of 3 het aantal responderende bedrijven kleiner is dan 7, voor gk 1, 2 en 3 samen een variantie bepaald.
- Voor de IS is als voor een kerncel in gk 4 het aantal responderende bedrijven kleiner is dan 7, bij de variantiebepaling gk 4 samengenomen met gk 5.
- Voor de PS vindt deze samenvoeging van gk 4 en 5 plaats als in ten minste één van deze twee grootteklassen het aantal responderende bedrijven kleiner is dan 7.

Evenals vorig jaar zijn gegevens van andere strata overgenomen om tot een schatting van stratumgemiddelden en -varianties te komen als bovengenoemde strategie niet werkte vanwege onvoldoende gesimuleerde responsdata. Concreet zijn de beperkte gegevens uit te mager gevulde strata samengevoegd met die uit verwante strata (tussen ConGO/niet-ConGO, met strata met gelijke eerste

3, 2, 1, digits qua kerncel/SBI-aanduiding) als dit niet lukte. Deze laatste stap zal i.h.a. een overschatting van de variantie geven, waarmee er een ingebouwde veiligheid is als op basis daarvan fracties verhoogd worden.

7 Steekproefontwerp NOPS 2009

7.1 Werkwijze

Zoals in het voorafgaande beschreven zijn op de drie relevante aggregatieniveaus de variatiecoëfficiënten bepaald. In de meeste gevallen werd met het oude ontwerp aan de geformuleerde kwaliteitseisen voldaan. Waar dit niet het geval was, zijn fracties handmatig verhoogd op stratumniveau om op het geaggregeerde niveau aan de eisen te voldoen.

7.2 Resultaten totaal PS en REGKOLs

Het resultaat voor de variatiecoëfficiënt voor toegevoegde waarde voor het totaal van de PS bedroeg 0,7% (de waarde voor omzet is minder bruikbaar omdat deze in het PS plus domein geïmputeerd zou moeten worden met (te) hoge varianties; voor ‘toegevoegde waarde’ is in het PS-plus domein in deze berekening BTW informatie voor ‘toegevoegde waarde’ gebruikt). Voor de REGKOLs lag in 80 gevallen de variatiecoëfficiënt voor de omzet (de natuurlijke variabele voor nationale rekeningen) onder de 5%, in 21 overige gevallen tussen 5% en 20% en in één geval op 48,2%. Dit laatste is implausibel en de variatiecoëfficiënt voor ‘toegevoegde waarde’ in deze laatste REGKOL bedroeg 1,7%, wat aannemelijker lijkt.

7.3 Herijking PS-plus

In overleg met BES is voor de allocatie voor herijking van de PS-plus het uitgangspunt gehanteerd dat voor de PS-plus gerelateerde strata dezelfde behandeling geldt als voor de niet PS-plus strata. De oude allocatie (Salemink e.a., 2001) voor de PS was een Neyman-allocatie die op detailniveau is aangepast om op lagere aggregatieniveaus (de kerncellen) goede cijfers te geven. Dat is ook voor de allocatie voor de herijking van de PS-plus gedaan. Het ontwerp voor de herijking is dan concreet te beschouwen als een ‘ingepaste’ Neyman-allocatie, d.w.z. ingepast in de al bestaande niet-PS-plus allocatie en vanuit hetzelfde concept als de niet-PS-plus allocatie. Expliciet betekent dit dat het steekproefaantal per PS-plus stratum h evenredig is met $N_h \hat{S}_h$ (Zie Cochran, p. 98). Binnen het PS-plus domein komt dit neer op een Neyman-allocatie binnen het hele PS-plus domein. De steekproefmassa per PS-plus stratum h wordt dan

verdeeld volgens:

$$n_h = n_{\text{herijking}} \frac{N_h \hat{S}_h}{\sum_{h \in \text{PS-plus strata}} N_h \hat{S}_h} \quad (9)$$

Hierin is $n_{\text{herijking}}$ de steekproefmassa die voor de herijking beschikbaar is. Er is een groot deel van de steekproefmassa toegekend via vgl. (9) door een lagere waarde (19.600) voor $n_{\text{herijking}}$ te nemen dan de beschikbare 20.000 eenheden. Zo blijft er massa over voor een evt. verfijning voor lagere niveaus (de kerncellen), waar nodig. Vervolgens is gekeken of met deze toekenning aan de algemene kwaliteitscriteria voor NOPS voldaan is (op totaal-PS-niveau, REGKOL- en kerncelniveau). Waar nodig heeft dit tot handmatige verhoging van de fracties geleid, onder gebruikmaking van het restant van de voor herijking beschikbare steekproefmassa. Zo wordt voldaan aan de richtlijnen uit Stroeks (2006).

Voor deze berekening is voor de waarden van \hat{S}_h de BTW-data voor ‘toegevoegde waarde’ binnen het PS-plus domein gebruikt. Dit model is opgesteld in overleg met BES en toegelicht bij de bespreking van de reacties op de eerste versie van het steekproefontwerp en ook nadien in gesprekken met inhoudelijk deskundigen van de PS-plus akkoord bevonden.

Het bleek dat voor de kerncellen met een PS-plus component aan de eis van een variatiecoëfficiënt kleiner dan 20% ruimschoots voldaan wordt (zoals overigens ook voor de meeste niet PS-plus kerncellen). De verdeling van de variatiecoëfficiënten laat voor de aan PS-plus strata gerelateerde kerncellen eenzelfde beeld zien als de niet aan PS-plus strata gerelateerde kerncellen.

Na afstemming met inhoudelijk deskundigen is in sommige strata de steekproefmassa verhoogd om op stratumniveau het ontwerp geschikter te maken voor de herijking wat betreft de structuren. Dit gaat boven de gestelde en in het verleden gedocumenteerde eisen uit, maar dit is onmiskenbaar nuttig en het belang van de geschiktheid voor de herijking van de structuren is ook vanuit DMH naar voren gebracht toen de eisen gesteld aan het steekproefontwerp voor de herijking vastgesteld werden.

7.4 Aanpassingen op verzoek van de statistiekafdelingen

Vanuit de statistiek-afdelingen zijn de volgende verzoeken tot aanpassing van de allocatie voor de PS gedaan:

- Voor de zorgpraktijken van fysiotherapeuten en verloskundigen (kerncellen 869120 resp. 869110) hoeft geen steekproef getrokken te worden. De fracties in de bijbehorende strata zijn op 0 gezet.
- In gk 1 van de industrie (kerncellen waarvan de codes beginnen met 10 tot en met 33, behalve 32991) wordt niet waargenomen. De fracties in de bijbehorende strata zijn op 0 gezet.

- In kerncel 68000x (verhuur en handel in onroerend goed) zijn bedrijven met SBI2008-codes SBI 68201 en SBI 68202 buiten de steekproeftrekking gehouden. Dit is geforceerd tijdens het steekproeftrekken. Ook zijn in deze kerncel handmatig 500 steekproefeenheden toegekend, er werd hier voor het eerst waargenomen. De steekproefmassa is in mindering gebracht op de steekproef binnen de agrarische dienstverlening, kerncel 01600 (waar de variatiecoëfficiënt een zeer lage waarde had). Hierbij is de resterende steekproefmassa via een Neyman-allocatie binnen het steekproefdeel herverdeeld.

7.5 Aanpassingen vanuit kwaliteitsoogpunt

Uit de variantieberekeningen is gebleken dat voor een aantal kerncellen de variatiecoëfficiënt de 20% overschreed. In die 11 gevallen zijn fracties handmatig verhoogd. Hierbij was ‘toegevoegde waarde’ leidend. Er zijn (afgezien van in kerncel 01600) geen aanpassingen gedaan die tot een hogere variatiecoëfficiënt hebben geleid.

7.6 Situaties waar de kwaliteitseisen niet gehaald zijn

Ondanks de aanpassingen is niet in alle gevallen de variatiecoëfficiënt per REGKOL onder de 5% gekomen. Voor verdere verhoging van de fracties is niet gekozen omdat de enquêtevakantie belangrijker werd gevonden. De overschrijding m.b.t. de eisen op de REGKOLs is in deze gevallen dus geaccepteerd; dit in overleg met projectleiding en senior users.

Ook voor een aantal kerncellen voldeden de variatiecoëfficiënten per kerncel uiteindelijk niet geheel aan de eisen. Nadere inspectie liet zien dat de imputatie (zie paragraaf 6.3) van de varianties in die gevallen tot (implausibel) hoge schattingen van de varianties had geleid; invullen van de stratumvarianties van vorig jaar in de berekening gaf in die gevallen een variatiecoëfficiënt die wel acceptabel is.

8 Steekproefontwerp IS2009

8.1 Aanpassingen vanuit kwaliteitsoogpunt

Ondanks de wens van een gelijke steekproef in vergelijking met statistiekjaar 2008 zijn een aantal fracties in het steekproefdeel aangepast wanneer de variatiecoëfficiënt een waarde had hoger dan 40%. Deze aanpassingen zijn handmatig aangebracht en vaak maar voor zover enquêtevakantie dit toeliet: de grenzen van de enquêtevakantie zijn per stratum als maximum gehanteerd.

9 De steekproeven en steekproefomvang

De steekproeven voor NOPS en IS zijn voorgelegd aan de senior users en andere inhoudelijk deskundigen en, na enkele verzoeken tot aanpassing die geïmplementeerd zijn, goedgekeurd door de senior users.

De steekproef voor NOPS omvatte uiteindelijk 95.969 eenheden, inclusief de 19339 eenheden voor de herijking van de PS-plus. Hieraan zijn de ruim drieduizend ConGO eenheden toegevoegd (inclusief eenheden uit gk 0 en mogelijk eenheden die niet tot de populatie van de PS en de IS horen).

De eerste steekproef voor de IS omvatte 54.118 eenheden. Hieraan zijn de ruim drieduizend ConGO eenheden toegevoegd (inclusief eenheden uit gk 0).

Zowel voor de PS als voor de IS is het totaal aantal steekprofeenheden kleiner dan voor statistiekjaar 2007 (voor de PS 80.549 enquêtes en voor de IS 60.542 enquêtes), waarmee qua aantallen aan de doelstellingen is voldaan.

Met het oog op enquêtevakantie is de steekproef voor de IS door afronding klein uitgevallen. Daarom is de steekproef voor de IS2009 recent opnieuw bekeken en o.m. vergeleken met die voor de IS2008 en de net beschikbaar gekomen respons van de IS2008. Van deze nadere inspectie is de conclusie dat de IS steekproef voor statistiekjaar 2009 vrijwel geheel (op twee kerncellen na) geschikt is.⁶ Voor kerncel 75000 voor de PS en IS blijkt ook waarneming nodig te zijn. Zekerheidshalve hebben inhoudelijk deskundigen van de PS nogmaals naar de PS steekproef gekeken. E.e.a. heeft tot aangepaste steekproefaantallen in zes kerncellen geleid die dan in een bijtrekking voor de PS en de IS kunnen worden gerealiseerd.

10 Conclusie

De impact van de beide veranderingen (overgang op BE_{nieuw} en integrale waarneming ConGO) in de PS en IS voor de steekproef is beperkt: de overgang op nieuwe statistische eenheid door de invoering van de OG+ leidt tot kleine veranderingen in het kader, met name in het steekproefdeel, vanwege het grote aantal 1-op-1 relaties tussen BE_{nieuw} en BE_{oud} . Ook de invloed van het integraal waarnemen van ConGO op de steekproefmassa is beperkt: het integraal waarnemen voor ConGO speelt zich deels (ongeveer 1400 eenheden) in het niet-steekproef deel af waar altijd al integrale waarneming was en het gaat om een beperkt aantal eenheden in het deel waar wel een steekproef is (1200 eenheden, sommige mogelijk buiten de PS en IS). In combinatie met de gebruikelijke

⁶Een ander resultaat van dit nadere onderzoek is dat voor de IS niet zozeer de variatiecoëfficiënt op kerncelniveau van belang is, maar die op het aggregatieniveau van 3-digit SBI, het publicatieniveau voor de IS. Dit kan voor volgend jaar een uitgangspunt zijn voor alternatieve kwaliteitscriteria voor de IS.

populatiodynamiek en economische ontwikkelingen heeft e.e.a. niet tot veel overschrijdingen van de variatiecoëfficiënten in relatie tot de kwaliteitseisen geleid. Daarom kan het steekproefontwerp voor statistiekjaar 2009 grotendeels identiek zijn aan dat voor statistiekjaar 2008, omdat dat aan de eisen voor dat jaar voldeed en de eisen voor 2009 niet veel afwijken van die voor 2008.

De 20.000 extra eenheden voor de herijking van de PS-plus zijn voldoende gebleken om een steekproef voor de PS-plus strata te ontwerpen zodat voor de aan PS-plus gerelateerde kerncellen aan de algemene kwaliteitseisen voldaan is. Hiermee kan aan de in het verleden geformuleerde eis voor een steekproef in een herijkingsjaar voldaan worden. Op het nog lagere niveau van PS-plus strata, was ruimte om hier fracties nog te verhogen om structuren beter te kunnen schatten.

282 4 nee 31 0.55 0.29 17 8 * 47520 3 ja 631 0.14 0.17 87 107 * 96010 4 nee 47 0.57 0.19 26 8
282 5 nee 39 0.76 1.00 29 39 * 47520 4 nee 265 0.41 0.35 108 92 * 96010 5 nee 30 0.57 1.00 17 30
282 6 nee 18 1.00 1.00 18 18 * 47520 5 nee 97 0.85 1.00 82 97 * 96010 6 nee 8 1.00 1.00 8 8
282 7 nee 15 1.00 1.00 15 15 * 47520 6 nee 19 1.00 1.00 19 19 * 96010 7 nee 2 1.00 1.00 2 2
282 8 nee 7 1.00 1.00 7 7 * 47520 7 nee 11 1.00 1.00 11 11 * 96010 8 nee 2 1.00 1.00 2 2
282 9 nee 0 1.00 1.00 0 0 * 47520 8 nee 3 1.00 1.00 3 3 * 96010 9 nee 2 1.00 1.00 2 2
2822 1 ja 86 0.03 0.00 2 0 * 47520 9 nee 2 1.00 1.00 2 2 * 96020 1 ja 29585 0.01 0.00 150 0
2822 2 ja 40 0.07 0.04 2 1 * 47530 1 ja 365 0.02 0.02 6 9 * 96020 2 ja 4832 0.01 0.03 62 144
2822 3 ja 28 0.76 0.11 21 3 * 47530 2 ja 530 0.02 0.00 11 0 * 96020 3 ja 1217 0.02 0.06 22 73
2822 4 nee 37 0.41 0.29 15 10 * 47530 3 ja 159 0.23 0.20 35 31 * 96020 4 nee 251 0.08 0.07 20 17
2822 5 nee 53 0.76 1.00 40 53 * 47530 4 nee 25 1.00 1.00 25 25 * 96020 5 nee 48 0.18 1.00 8 48
2822 6 nee 24 1.00 1.00 24 24 * 47530 5 nee 4 1.00 1.00 4 4 * 96020 6 nee 3 1.00 1.00 3 3
2822 7 nee 16 1.00 1.00 16 16 * 47530 6 nee 3 1.00 1.00 3 3 * 96020 7 nee 2 1.00 1.00 2 2
2822 8 nee 7 1.00 1.00 7 7 * 47530 7 nee 0 1.00 1.00 0 0 * 96020 8 nee 1 1.00 1.00 1 1
2822 9 nee 2 1.00 1.00 2 2 * 47530 8 nee 3 1.00 1.00 3 3 * 96020 9 nee 0 1.00 1.00 0 0
2829 1 ja 176 0.02 0.00 3 0 * 47530 9 nee 0 1.00 1.00 0 0 * 96030 1 ja 404 0.01 0.00 4 0
2829 2 ja 99 0.08 0.04 7 3 * 47540 1 ja 341 0.02 0.15 7 51 * 96030 2 ja 279 0.07 0.09 19 25
2829 3 ja 75 0.13 0.11 9 8 * 47540 2 ja 315 0.10 0.10 30 31 * 96030 3 ja 85 0.25 0.13 21 11
2829 4 nee 80 0.18 0.29 14 23 * 47540 3 ja 51 0.17 0.03 8 1 * 96030 4 nee 54 0.65 0.20 35 10
2829 5 nee 69 0.76 1.00 52 69 * 47540 4 nee 12 0.26 0.29 3 3 * 96030 5 nee 25 0.62 1.00 15 25
2829 6 nee 25 1.00 1.00 25 25 * 47540 5 nee 0 0.63 1.00 0 0 * 96030 6 nee 5 1.00 1.00 5 5
2829 7 nee 17 1.00 1.00 17 17 * 47540 6 nee 0 1.00 1.00 0 0 * 96030 7 nee 0 1.00 1.00 0 0
2829 8 nee 3 1.00 1.00 3 3 * 47540 7 nee 0 1.00 1.00 0 0 * 96030 8 nee 1 1.00 1.00 1 1
2829 9 nee 0 1.00 1.00 0 0 * 47540 8 nee 0 1.00 1.00 0 0 * 96030 9 nee 3 1.00 1.00 3 3
2830 1 ja 77 0.00 0.00 0 0 * 47540 9 nee 0 1.00 1.00 0 0 * 96040X 1 ja 6991 0.10 0.00 699 0
2830 2 ja 65 0.10 0.04 6 2 * 47590 1 ja 2079 0.01 0.00 20 0 * 96040X 2 ja 1132 0.10 0.17 113 192
2830 3 ja 31 0.19 0.11 6 3 * 47590 2 ja 2219 0.02 0.05 50 104 * 96040X 3 ja 145 0.10 0.23 14 33
2830 4 nee 37 0.55 0.29 20 10 * 47590 3 ja 619 0.03 0.11 20 67 * 96040X 4 nee 63 0.30 0.24 18 15
2830 5 nee 37 0.76 1.00 28 37 * 47590 4 nee 176 0.26 0.29 45 51 * 96040X 5 nee 33 0.70 1.00 23 33
2830 6 nee 16 1.00 1.00 16 16 * 47590 5 nee 65 0.63 1.00 40 65 * 96040X 6 nee 9 1.00 1.00 9 9
2830 7 nee 9 1.00 1.00 9 9 * 47590 6 nee 20 1.00 1.00 20 20 * 96040X 7 nee 3 1.00 1.00 3 3
2830 8 nee 5 1.00 1.00 5 5 * 47590 7 nee 6 1.00 1.00 6 6 * 96040X 8 nee 0 1.00 1.00 0 0
2830 9 nee 0 1.00 1.00 0 0 * 47590 8 nee 3 1.00 1.00 3 3 * 96040X 9 nee 0 1.00 1.00 0 0
47590 9 nee 3 1.00 1.00 3 3

Appendix B: Inventarisatie opsplitsgraad nieuwe-oude BE

Zoals in §2 beschreven is elke BE_{nieuw} opgebouwd uit één of meer BE_{oud} -en. Het is instructief te weten wat de verdeling is van uit hoeveel BE_{oud} -en de BE_{nieuw} -en zijn opgebouwd. Een overzicht is te vinden in onderstaande tabel, gebaseerd op een prototype van EHB.

Uit deze tabel kunnen we concluderen dat in verreweg de meeste gevallen (96,4 %) een BE_{nieuw} is opgebouwd uit één BE_{oud} .

Tabel 1: Voor elke opsplitsgraad (aantal BE_{oud} per BE_{nieuw}) staat aangegeven hoeveel BE_{nieuw} deze opsplitsgraad hebben, en voor welke fractie van de BE_{nieuw} -en deze geldt.

aantal BE_{oud} per BE_{nieuw}	aantal BE_{nieuw}	fractie
1	1013282	0,964
2	22172	0,021
3	8825	0,008
4	3224	0,003
5	1453	0,001
6	708	0,001
7	326	0,000
8	220	0,000
9	121	0,000
10	71	0,000
11	52	0,000
12	43	0,000
13	18	0,000
14	18	0,000
15	15	0,000
16	8	0,000
17	5	0,000
18	6	0,000
19	6	0,000
20	4	0,000
21	1	0,000
22	1	0,000
24	1	0,000
26	1	0,000
27	1	0,000
30	1	0,000
31	1	0,000
33	1	0,000
35	1	0,000
38	1	0,000
40	1	0,000
43	1	0,000
64	1	0,000
97	1	0,000
102	1	0,000
107	1	0,000
112	1	0,000
som		1,000

Bibliografie

- Aelen, Frank (2009), *HEcS+. Methodologie Nieuwe Opzet Productiestatistieken (Versie 1.0p1)*, Interne nota, Centraal Bureau voor de Statistiek.
- Aelen, Frank en Carlo Driesen (2009), *NOPS Project Initiatie Document (Versie 1.0p2)*, Centraal Bureau voor de Statistiek.
- Beuken, Ivo (2010), *Eenhedenbase (EHB) Aanpassingen in ABR (Versie 1.0p22)*. Interne nota, Centraal Bureau voor de Statistiek, Heerlen.
- Brocke, P. (2004), *Onderzoeksbeschrijving Statistiek Investerings en lease (Versie 3.0)*. Interne nota BWV, Centraal Bureau voor de Statistiek, Voorburg.
- Van de Burgt, Guus (2006), *Draaiboek FESS, Deel II*,
\\Besh1f\Diensten1\G0004\ Postmap\PsPlus\Documentatie
- Cochran, William G., *Sampling Techniques* (1977). Wiley, New York.
- Lammertsma, Alex (red.) (2010), *Business Architectuur Document NOPS 2009 (Versie 0.4p2)*. Interne nota. Centraal Bureau voor de Statistiek, Den Haag/Heerlen.
- Resing, Ben, Rik van der Heijden en Daisy Debie (2005). *Handboek Impect 1 PS-en. (Versie 1.0)*. Interne nota BOO, Centraal Bureau voor de Statistiek, Heerlen.
- Salemink, Irene, Pieter Vlag en Nico Nieuwenbroek (2001). *Steekproefallocatie Productie Statistieken 2000*. CBS-rapport, BPA-nr H1762-01-BOO, Centraal Bureau voor de Statistiek, Heerlen.
- Smeets, Marc, Thomas Kraan, Judit Arends-Tóth (2008), *Steekproefontwerp PS en IS 2007 met enquêtevakantie*. CBS-nota, BPA-nr DMH-2008-30-01-MSET, Centraal Bureau voor de Statistiek, Heerlen.
- Smeets, Marc, Thomas Kraan, Judit Arends-Tóth (2009a), *Steekproefontwerp PS en IS 2008*. CBS-nota, BPA-nr DMH-2009-02-26-MSET, Centraal Bureau voor de Statistiek, Heerlen.
- Smeets, Marc, Thomas Kraan, Judit Arends-Tóth (2009b), *Sampling design for the Dutch Structural Business Survey and the Dutch Investment Survey*. CBS-nota, BPA-nr DMH-04-06-09-MSET, Centraal Bureau voor de Statistiek, Heerlen.
- Stroeks, Leon (2006), *Draaiboek Structuren in PS plus, Deel III*,
\\Besh1f\Diensten1\G0004\ Postmap\PsPlus\Documentatie

definitie REGKOL. intranet site. <http://cbsh1sps/sites/ketenregie/Shared Documents/Hoofdenoverleg 5 juni 2009/4b. Nieuwe REGKOL-indeling NR.doc>